APPLICATION FOR UNITED STATES LETTERS PATENT

FOR

**Method and System to Pre-Fetch a Protocol Control Block
for Network Packet Processing**

Inventor: **Rameshkumar G. Illikkal**

Prepared by:

BLAKELY SOKOLOFF TAYLOR & ZAFMAN, LLP
12400 Wilshire Boulevard, 7th Floor
Los Angeles, California 90025
(503) 684-6200

Express Mail No. EV325525869US

Method and System to Pre-Fetch a Protocol Control Block

for Network Packet Processing

BACKGROUND

5 1.   Technical Field

[0001]   Embodiments of the invention relate to the field of network packet

processing, and more specifically to pre-fetching a protocol control block for network

packet processing.

10 2.   Background Information and Description of Related Art

[0002]   When a packet arrives at a network device, the network interface card

(NIC) takes the packet and stores it in a main memory. The NIC may then send an

interrupt to notify the central processing unit (CPU) about the packet. An interrupt

unit may then check the destination of the interrupt, disable further interrupts from

15 the NIC, initiate a software interrupt, and queue the packet for processing. When

the processing unit is ready to process the packet, the connection to which the

packet belongs is identified. This may involve fetching a Protocol Control Block

(PCB) associated with the packet. After the PCB is fetched, the CPU may start

processing the packet. The memory latency that occurs from fetching the PCB

20 when the processing unit is ready to process the packet decreases the performance

of the network device. As networking speeds increase, this memory latency

becomes an increasing problem for performance and throughput.

BRIEF DESCRIPTION OF DRAWINGS

[0003]   The invention may best be understood by referring to the following description and accompanying drawings that are used to illustrate embodiments of the invention. In the drawings:

[0004]   **FIG. 1** is a block diagram illustrating one generalized embodiment of a system incorporating the invention.

[0005]   **FIG. 2** is a block diagram illustrating an exemplary system incorporating the invention according to one embodiment of the invention.

[0006]   **FIG. 3** is a flow diagram illustrating a method according to an embodiment of the invention.

[0007]   **FIG. 4** is a block diagram illustrating a suitable computing environment in which certain aspects of the illustrated invention may be practiced.

DETAILED DESCRIPTION

[0008] Embodiments of a system and method to pre-fetch a protocol control block for network packet processing are described. In the following description, numerous specific details are set forth. However, it is understood that embodiments

5 of the invention may be practiced without these specific details. In other instances, well-known circuits, structures and techniques have not been shown in detail in order not to obscure the understanding of this description.

[0009] Reference throughout this specification to "one embodiment" or "an embodiment" means that a particular feature, structure, or characteristic described

10 in connection with the embodiment is included in at least one embodiment of the invention. Thus, the appearances of the phrases "in one embodiment" or "in an embodiment" in various places throughout this specification are not necessarily all referring to the same embodiment. Furthermore, the particular features, structures, or characteristics may be combined in any suitable manner in one or more

15 embodiments.

[0010] Referring to Fig. 1, a block diagram illustrates a system 100 according to one embodiment of the invention. Those of ordinary skill in the art will appreciate that the system 100 may include more components than those shown in Fig. 1. However, it is not necessary that all of these generally conventional components be

20 shown in order to disclose an illustrative embodiment for practicing the invention.

[0011] System 100 includes a receive unit 102 to receive packets from a network. In one embodiment, the packets may be communicated in accordance with the Transmission Control Protocol (TCP/IP) specification. A particular series of

packets may be referred to as a "connection" or "packet flow." The context of a connection may be stored in a structure known as a Protocol Control Block (PCB). This context may be uniquely identified by the connection's source IP address, destination IP address, source port, destination port, and/or protocol type. For each

5   packet received at receive unit 102, the PCB associated with the packet may need to be retrieved from memory.

[0012]   Accessing the PCB from the memory for each packet sent and received has associated memory latency and bandwidth issues. To reduce these issues, a pre-fetch unit 104 fetches the PCB associated with a received packet into a cache

10   108 of a processing unit 106. The packet is then queued for processing. When the processing unit 106 is ready to process the packet, the PCB may be retrieved from its cache 108. Thus, the memory latency in fetching the PCB when the processing unit is ready to process the packet is reduced.

[0013]   In one embodiment, the pre-fetch unit 104 also pre-fetches packet header

15   information into the cache 108. When the processing unit 106 is ready to process the packet, the packet's header information may be retrieved from the cache 108 and the packet may then be processed.

[0014]   Fig. 2 illustrates an exemplary system incorporating the invention according to one embodiment of the invention. In this embodiment, a network

20   interface card (NIC) 202 receives packets from a network. The packet is stored in a main memory 214 through memory controller 212. The NIC 202 sends an interrupt to notify a processing unit, such as 206, 208, or 210, about the packet. An interrupt unit 204, such as an interrupt service rotate (ISR) unit, checks the destination of the

interrupt, disables further interrupts from the NIC, initiates a software interrupt, and queues the packet for processing. In one embodiment, the packet is queued for processing by queuing a deferred procedure call (DPC). At this time, a pre-fetch of the PCB associated with the packet may be initiated. A pre-fetch of a packet header

5   may also be initiated. The PCB and packet header may be pre-fetched into a cache, such as 216, 218, or 220. The pre-fetch may be done in hardware or software. When a processing unit, such as 206, 208, or 210, is ready to process the packet, the processing unit may fetch the PCB and packet header from its cache. Then, the packet may be processed. The processing unit may then enable

10   interrupts from the NIC.

[0015]   In one embodiment, pre-fetching may also be used on the send side to reduce memory latency. When a packet is queued for transmission out of the network, the PCB may be pre-fetched. There is usually some delay between the socket interface in the user space and the protocol stack processing in the kernel

15   space. Therefore, when a send request is initiated for a packet, the PCB associated with the packet may be pre-fetched. When the kernel is ready to process the packet for transmission, the PCB has already been pre-fetched and is ready for processing. This reduces the memory latency on the send side.

[0016]   Fig. 3 illustrates a method according to one embodiment of the invention.

20   At 300, a packet is received. At 302, a PCB associated with the packet is pre-fetched into a cache. In one embodiment, a packet header is also pre-fetched into the cache. At 304, the packet is queued for processing. At 306, when a processing unit is ready to process the packet, the PCB is retrieved from the cache. In one

embodiment, the packet's header is also retrieved from the cache. Then, the packet may be processed.

[0017]   Fig. 4 is a block diagram illustrating a suitable computing environment in which certain aspects of the illustrated invention may be practiced. In one

5   embodiment, the method described above may be implemented on a computer system 400 having components 402 – 412, including a processor 402, a memory 404, an Input/Output device 406, a data storage device 412, and a network interface 410, coupled to each other via a bus 408. The components perform their conventional functions known in the art and provide the means for implementing the

10   system 100. Collectively, these components represent a broad category of hardware systems, including but not limited to general purpose computer systems and specialized packet forwarding devices. It is to be appreciated that various components of computer system 400 may be rearranged, and that certain implementations of the present invention may not require nor include all of the

15   above components. Furthermore, additional components may be included in system 400, such as additional processors (e.g., a digital signal processor), storage devices, memories, and network or communication interfaces.

[0018]   As will be appreciated by those skilled in the art, the content for implementing an embodiment of the method of the invention, for example, computer

20   program instructions, may be provided by any machine-readable media which can store data that is accessible by system 100, as part of or in addition to memory, including but not limited to cartridges, magnetic cassettes, flash memory cards, digital video disks, random access memories (RAMs), read-only memories (ROMs),

and the like. In this regard, the system 100 is equipped to communicate with such machine-readable media in a manner well-known in the art.

[0019]   It will be further appreciated by those skilled in the art that the content for implementing an embodiment of the method of the invention may be provided to the

5   system 100 from any external device capable of storing the content and communicating the content to the system 100. For example, in one embodiment of the invention, the system 100 may be connected to a network, and the content may be stored on any device in the network.

[0020]   While the invention has been described in terms of several embodiments,

10   those of ordinary skill in the art will recognize that the invention is not limited to the embodiments described, but can be practiced with modification and alteration within the spirit and scope of the appended claims. The description is thus to be regarded as illustrative instead of limiting.

_____

15

Express Mail No. EV325525869US